Volume 282, number 1, 41–46 FEBS 09614 April 1991

ADONIS 001457939100336

The 'positive-inside rule' applies to thylakoid membrane proteins

Ylva Gavel[1,3], Johannes Steppuhn[3], Reinhold Herrmann[3] and Gunnar von Heijne[2]

[1] *Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden,* [2] *Department of Molecular Biology, Karolinska Institute, Center for Biotechnology, NOVUM, S-141 57 Huddinge, Sweden* and [3] *Botanisches Institut der Ludwig-Maximilians-Universität, Menzinger Strasse 67, D-8000 München 19, Germany*

Received 11 February 1991

In integral membrane proteins, regions that span the lipid bilayer alternate with regions that are exposed on either side of the membrane. For proteins from the plasma membrane of both prokaryotic and eukaryotic cells it has been shown that the exposed parts follow a 'positive-inside rule': on average, segments that are translocated across the membrane have a 2–4-fold lower frequency of positively charged residues than non-translocated segments. We now present an analysis of proteins from the thylakoid membrane of chloroplasts. It is shown that these proteins have the same charge asymmetry as has been reported for proteins from other membrane systems, with their more highly charged regions facing the stromal compartment.

Thylakoid; Membrane protein; Topology; Positive inside rule

1. INTRODUCTION

The polypeptide backbone of most integral membrane proteins has at least one long hydrophobic segment that spans the membrane, most likely as an α-helix [1]. This applies to proteins from the plasma membrane of both prokaryotes and eukaryotes; however, proteins from the outer membrane of Gram-negative bacteria are built on a different set of structural principles, and will not be considered further here [2]. Many integral membrane proteins are polytopic, i.e. have several membrane-spanning segments connected by surface-exposed loops. The topology is a specification of the membrane-embedded domains and the sidedness of the protein.

Proteins from different membrane systems appear to have certain topogenic signals in common: statistical studies correlating the transmembrane topology with the amino acid sequence indicate that proteins from the bacterial inner membrane as well as proteins from the plasma membrane of eukaryotes follow a 'positive-inside rule' [3,4]. This conclusion was originally based on the observation that positively charged residues (i.e. Lys and Arg) tend to be enriched in non-translocated (i.e. cytoplasmically exposed) as compared to translocated segments of the protein. Recently, substantial experimental support for this notion has been reported [5–13]. These results suggest that it is possible to predict the transmembrane orientation of a protein once the putative membrane-spanning segments have been identified by hydrophobicity analysis: a topology that is consistent with the positive-inside rule should be most probable.

The evidence that the topology of proteins from the bacterial inner membrane and membrane proteins in the secretory pathway of eukaryotes follow the positive-inside rule is thus quite solid. However, for lack of sequence and topology data, it has hitherto not been possible to extend the analysis to include proteins of the thylakoid membrane of chloroplasts. In this paper, we analyse the topological distribution of amino acids in a number of recently published chloroplast protein sequences where all or parts of the topology is known from experimental data, and show that the positive-inside rule applies also to this class of proteins.

Correspondence address: G. von Heijne, Department of Molecular Biology, Karolinska Institute, Center for Biotechnology, NOVUM, S-141 57 Huddinge, Sweden

2. MATERIALS AND METHODS

2.1. *Sequence collection and analysis*

Amino acid sequences of intrinsic thylakoid membrane proteins were collected from the literature. In cases where several closely related proteins were found, only one of them was included in the collection. Hydrophobicity analysis of the sequences was performed using a 19-residue moving window and the Engelman–Steitz hydrophobicity scale [14] as described previously [3,4].

We found 10 proteins for which experimental data on the transmembrane topology are available (Table I). In most cases, the mappings have been made by proteolytic digestion and/or antibody binding assays. Some proteins possess phosphorylation sites, and, since ATP does not readily cross the thylakoid membrane, the phosphorylated residues should be exposed to the stroma, where ATP is available [15]. Furthermore, a number of proteins are homologous to bacterial membrane proteins with known topology.

In some cases, the transmembrane topology is only partially known from experimental data. We have tried to extend the topology of these proteins using hydrophobicity analysis as described [3,4]. This requires that the sequence neighbouring the mapped parts contains segments hydrophobic enough (mean hydrophobicity ≥ 1.5) to be unambiguously assigned as putative membrane-spanning helices alternating with segments hydrophilic enough (mean hydrophobicity

Table I

Topologies of thylakoid membrane proteins. Residue no. 1 corresponds to the N-terminus of the mature protein. For experimental evidence: see footnotes

Protein source	Reference	Transmembrane topology		
		Stroma	Thylakoid membrane	Lumen
a. Nuclear-encoded proteins without lumen-targeting signals				
Major chlorophyll *a/b*-binding protein (CAB), pea[1]	[23–29]	1–67[3]	68–86[3]	87–118
		138–186	119–137[3]	
			187–205[3]	
				206–233[4]
b. Nuclear-encoded proteins with lumen-targeting signals				
CF_0 subunit II, spinach[5]	[30]			1–10
		147[6]	11–29	
c. Plastid-encoded proteins without presequences				
10 kDa phospho-protein, spinach[7]	[15,31,32]	1–39[8]	40–58	59–73
CF_0 subunit III, spinach	[21,22,33,34]			1–3[9]
		23–62	4–22	
			63–81	
				81–81[9]
Cytochrome *b*-559 α-subunit, spinach[7]	[35–38]	1–24	25–43	44–83[10]
Cytochrome *b*6, spinach	[39,40]	1–33[11]	34–52	53–87
		107–111	88–106	
			112–130	131–183[12]
		203–211[13]	184–202	
D1 protein, spinach[7]	[32,41,42]	1–34[8,14]	35–53[15]	54–108[16,17]
		128–141[18]	109–127[15]	
			142–160[15]	161–195[19]
		215–272[20,21,22]	196–214[15]	
			273–291[15]	292–353[23]
44 kDa chlorophyll *a* apoprotein, spinach[24]	[32,43]	1–48[25]	49–67	68–84
		104–111	85–103	
			112–130	131–158
			159–177*	
Cytochrome *f*, pea	[44,45]		252–270	1–251[27]
		271–285[26]		
CF_0 subunit I, spinach	[21,22,46,47]		9–27	1–8[28]
		28–166[28]		

Footnotes to Table I

[a] The topology could not be extended beyond this point
[b] Some of the experimental data have been obtained for CAB proteins from other species, but the topology of CAB proteins from higher plants is probably universal
[c] Proteolytic digestion experiments and antibody binding assays indicate that the N-terminus is located in the stroma. In accordance with this, a threonine residue in this region is phosphorylated
[d] It has recently been shown that CAB has 3 membrane spans (W. Kühlbrandt, personal communication)
[e] Since there are 3 membrane spans and the N-terminus resides in the stroma, the C-terminus should be located in the lumen. Proteolytic digestion experiments with carboxypeptidase A are consistent with this. Furthermore, surface radioiodination experiments indicate that Tyr-225 is exposed at the luminal surface
[f] This protein has recently been sequenced (Steppuhn et al., in preparation)
[g] Proteolytic digestion experiments indicate that a large portion of this protein is exposed to the stroma. Since the only transmembrane arrangement that would be consistent with the hydrophobicity profile is a topology with one membrane span near the N-terminus, the exposed part is probably the region between the putative membrane span and the C-terminus
[h] Including the initiator Met, which is posttranslationally cleaved
[i] Thr-2 is phosphorylated, indicating that the N-terminus is exposed to the stroma
[j] The sequence of this protein is 30% homologous to the sequence of CF_0 subunit *c* of *E. coli*. In subunit *c*, the N-terminus and the C-terminus are translocated across the membrane and face the periplasm
[10] When thylakoid membranes are exposed to trypsin or V8 protease, part of the protein is digested without loss of reactivity to an antibody to a C-terminal epitope. When inside-out vesicles are exposed to trypsin, reactivity to the antibody to the C-terminal epitope is lost. Immunogold labeling also suggests that the C-terminus is exposed to the lumen
[11] Trypsin sensitivity of an N-terminal epitope indicates that the N-terminus is exposed to the stroma
[12] A luminal orientation of this segment would account for the presence of trypsin-sensitive sites on the lumen side
[13] Trypsin sensitivity of a C-terminal epitope indicates that the C-terminus is exposed to the stroma
[14] An antibody to residues 22–31 binds to the stromal surface
[15] In addition to the mapping with antibodies, the topology is consistent with proteolytic digestion experiments, sequence homology to the *Rhodopseudomonas* reaction center and exon-intron boundaries in the D1 genes of *Chlamydomonas* and *Euglena*
[16] Increased binding of an antibody to residues 59–68 was observed for inside-out vesicles after washing, suggesting a luminal location of this epitope
[17] Increased binding of an antibody to residues 97–105 was observed for inside-out vesicles after washing, suggesting a luminal location of this epitope
[18] An antibody to residues 127–136 binds to the stromal surface
[19] An antibody to residues 165–174 binds to the luminal surface
[20] An antibody to residues 225–235 binds to the stromal surface
[21] An antibody to residues 261–270 binds to the stromal surface
[22] Point mutations at Val-219, Phe-255 and Ser-264 confer herbicide resistance. This suggests that these residues are located on the stromal side
[23] An antibody to residues 301–310 binds to the luminal surface
[24] Including the first 14 amino acids, which are trimmed off but do not look like a lumen targeting domain
[25] Thr-15 is phosphorylated, indicating that this residue is exposed to the stroma
[26] Digestion with carboxypeptidase A indicates that the C-terminus is exposed on the stromal side. Proteolytic digestion experiments indicate that the exposed sequence is rather short
[27] There is a putative heme-binding site near the N-terminus. The heme is thought to interact with plastocyanin, a luminal protein. Incubation of disrupted membranes or inside-out vesicles with proteinase K results in loss of heme-peroxidase activity. Activity is not affected in right-side-out vesicles or intact membranes
[28] The sequence of this protein is 19% homologous to the sequence of CF_0 subunit *b* of *E. coli*. In subunit *b*, the N-terminus is translocated across the membrane, whereas a large C-terminal region remains in the cytoplasm

≤ 0.7) to be unambiguously assigned as extramembraneous. Sequence stretches containing hydrophobic peaks in the ambiguous interval (i.e. mean hydrophobicity between 0.8 and 1.4) cannot be analyzed by this method, since a safe prediction of the number of membrane spans is not possible.

Our dataset was prepared using the sequences of the well-characterized proteins described above. Those parts of the proteins where the topology had not been mapped by experiments or where it could not be reliably inferred from hydrophobicity analysis were omitted. The topologies are shown in Table I.

From this dataset, 4 smaller samples were prepared: membrane-spanning segments with the N-terminus oriented towards the stroma, membrane-spanning segments with the N-terminus oriented towards the lumen, exposed segments facing the stroma, and exposed segments facing the lumen. The membrane-spanning parts were assumed to encompass the most hydrophobic 19-residue stretch of

each candidate transmembrane region. Previous studies indicate that only short exposed segments have a bias in the amino acid distribution, with the transition region somewhere between 70 and 100 residues [4]. Therefore, luminal and stromal segments >70 residues in length were excluded from the analysis.

2.2. *Statistical analysis*

The amino acid frequencies were calculated for each subsample. The statistical significance of differences between the amino acid frequencies in different samples was estimated by χ^2-analysis.

3. RESULTS

We have analyzed the amino acid composition of the transmembrane and surface-exposed luminal and stromal domains of 10 non-homologous thylakoid membrane proteins with partly or fully determined topologies, Table I. In particular, the distribution of charged amino acids between the luminal and stromal domains has been calculated, since we wanted to find out whether thylakoid membrane proteins conform to the same 'positive-inside rule' that is known to apply to membrane proteins from other membrane systems.

3.1. *The amino acid composition of the transmembrane segments is similar to that of prokaryotic and eukaryotic membrane proteins*

The overall amino acid composition of the sample of transmembrane segments from the multi-spanning thylakoid membrane proteins of Table I is shown in Fig. 1a. There are no significant differences between the thylakoid sample and previously analyzed samples of prokaryotic and eukaryotic plasma membrane proteins. On the other hand, although the sample is small, the amino acid composition of single-spanning thylakoid membrane proteins appears more similar to that observed for the corresponding group of eukaryotic proteins than for the prokaryotic ones, with a rather low frequency of alanine (~10%) and a high frequency of leucine (~25%, Fig. 1b). Transmembrane segments with the N-terminus oriented towards the stroma do not differ significantly from those with the N-terminus oriented towards the lumen (data not shown).

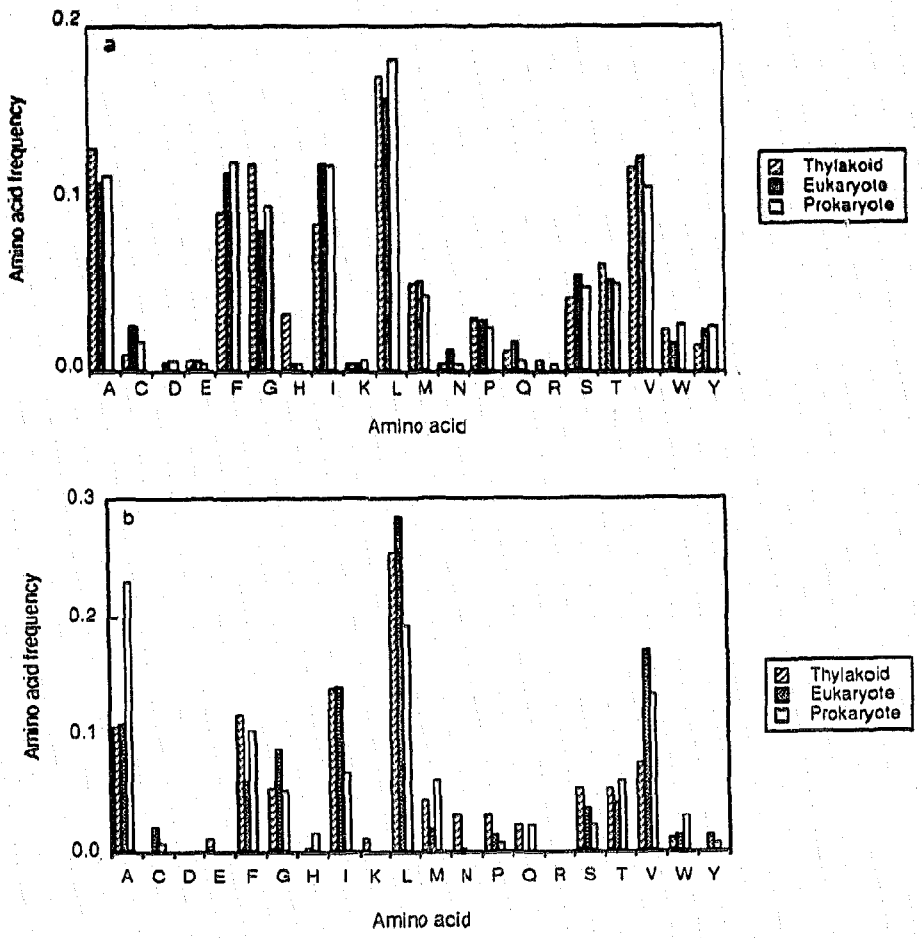

Fig. 1a. Amino acid frequencies in the transmembrane segments of multi-spanning membrane proteins. The frequencies for prokaryotic and eukaryotic proteins are from reference [4].

Fig. 1b. Amino acid frequencies in the transmembrane segments of single-spanning membrane proteins. The frequencies for prokaryotic and eukaryotic proteins are from reference [4].

3.2. *The amino acid composition of the luminal and stromal domains follows the 'positive-inside rule'*

The amino acid composition of exposed segments facing the stroma differs significantly from that of exposed segments facing the lumen (Table II). In particular, the number of positively charged residues (Arg and Lys) is enriched more than two-fold in stromal as compared to luminal loops (12.9% vs 5.4%, $P < 3 \cdot 10^{-4}$). Taken individually, both Arg (7.5% vs. 3.6%, $P < 0.02$) and Lys (5.4% vs. 1.9%, $P < 0.01$) were significantly enriched in the stromal loops. No similar bias in the distribution of negatively charged residues is found, with the frequency of Asp + Glu being 10.4% in the stromal loops and 10.9% in the luminal loops. This is similar to what has previously been found for both bacterial and eukaryotic plasma membrane proteins.

4. DISCUSSION

The results presented above demonstrate that the distribution of positively charged residues in the exposed parts of thylakoid membrane proteins has a sidedness of the same type as has been reported previously for other membrane systems [3,4]: Arg and Lys are less abundant in the luminal segments that have been translocated across the membrane than in the stromal segments that remain non-translocated. However, the bias is not as striking as in bacterial inner membrane proteins, where the difference in the frequency of Arg and Lys is almost four-fold [3]. Still, the strict alternation of the absolute number of positive charges in translocated vs. non-translocated loops, which is often found in the latter group, can be seen in most thylakoid proteins as well. We also note that negatively charged residues are evenly distributed between luminal and stromal domains, again as has previously been observed for other membrane proteins. Indeed, negatively charged residues have recently been shown to be much less potent topological determinants than positively charged ones for *E. coli* membrane proteins [7].

Table II

Amino acid frequencies in exposed loops

Amino acid	Stromal	Luminal
A	0.06576	0.07126
C	0.00454	0.00000
D	0.02494	0.04751
E	0.07937	0.06176
F	0.04989	0.06888
G	0.10884	0.08551
H	0.01134	0.02375
I	0.04535	0.06651
K*	0.05442*	0.01900*
L	0.08390	0.07601
M	0.01814	0.02613
N	0.03628	0.07126
P	0.05442	0.06176
Q	0.02268	0.02613
R*	0.07483*	0.03563*
S	0.08844	0.08789
T	0.06576	0.04038
V	0.04762	0.06413
W	0.02268	0.02375
Y	0.04082	0.04276

Amino acids that differ significantly in frequency between stromal and luminal loops are marked with an asterisk

For newly sequenced thylakoid membrane proteins where the topology cannot be inferred from experimental data, our results suggest two complementing strategies for predicting the sidedness once the hydrophobicity profile has been determined. First, if the protein has a presequence containing a thylakoid-targeting signal (i.e. a signal-peptide like segment, [16]) it is highly likely that the N-terminus of the mature protein will be luminally exposed since thylakoid-targeting signals are cleaved by a luminally disposed protease [17–19]. Second, the positive-inside rule can be applied to predict the topology of the entire molecule, whether or not there is a thylakoid-targeting presequence. The topology that minimizes the number of positively charged residues (Arg and Lys) in the lumen should be chosen (note that the positive-inside rule does not apply when a translocated segment is very long, since only segments shorter than about 70 residues appear to have a reduced content of arginines and lysines [3,4].

If the protein lacks a thylakoid-targeting sequence, only the positive-inside rule can be used as a guide. There are many bacterial and eukaryotic membrane proteins that are made without a cleavable N-terminal signal peptide, yet assemble into the cytoplasmic membrane with the N-terminus translocating to the extracytoplasmic side [3,4,20]. Among the thylakoid membrane proteins, CF_0 subunit III seems to provide such an example: it lacks a thylakoid-targeting sequence but its topology as predicted by homology to its bacterial relative is nevertheless one with the N-terminus protruding into the lumen [21,22], which is also consistent with the positive-inside rule. When analyzing thylakoid membrane proteins lacking a thylakoid-targeting signal and with unknown topology, we have again found examples where the positive-inside rule predicts a luminally exposed N-terminus, and other examples where the opposite topology is predicted. For example, for subunit IX of photosystem I [48] we predict a translocated N-terminus and membrane spans between the residues 3–21 and 63–81. On the other hand, the β-subunit of cytochrome *b*-559 [36,38,49] (predicted to have a membrane span between the residues 19–37) should have its N-terminus exposed to the stroma according to the positive-inside rule.

In summary, we have shown that the positive-inside rule holds for all integral membrane proteins (except for bacterial outer membrane proteins, cf. section 1) analyzed so far, whether of prokaryotic, eukaryotic, or

organellar origin. This suggests that the mechanism of membrane assembly for this very important class of proteins is fundamentally similar throughout the living world.

Acknowledgements: This work was supported by a grant from the Swedish Natural Sciences Research Council to G.v.H., and grants from the German Research Foundation (SFB 184) and from the Fonds der Chemischen Industrie to R.H.

REFERENCES

[1] Jennings, M.L. (1989) Annu. Rev. Biochem. 58, 999–1027.
[2] Von Heijne, G. and Manoil, C. (1990) Prot. Engineer., in press.
[3] Von Heijne, G. (1986) EMBO J. 5, 3021–3027.
[4] Von Heijne, G. and Gavel, Y. (1988) Eur. J. Biochem. 174, 671–678.
[5] Von Heijne, G., Wickner, W. and Dalbey, R.E. (1988) Proc. Natl. Acad. Sci. USA 85, 3363–3366.
[6] Von Heijne, G. (1989) 341, 456–458.
[7] Nilsson, I. and Von Heijne, G. (1990) 62, 1135–1141.
[8] Haeuptle, M.T., Flint, N., Gough, N.M. and Dobberstein, B. (1989) J. Cell Biol. 108, 1227–1236.
[9] Parks, G.D., Hull, J.D. and Lamb, R.A. (1989) J. Cell Biol. 109, 2023–2032.
[10] Szczesna-Skorupa, E. and Kemper, B. (1989) J. Cell Biol. 108, 1237–1243.
[11] Sato, T., Sakaguchi, M., Mihara, K. and Omura, T. (1990) EMBO J. 9, 2391–2397.
[12] Boyd, D. and Beckwith, J. (1989) Proc. Natl. Acad. Sci. USA 86, 9446–9450.
[13] Yamane, K., Akiyama, Y., Ito, K. and Mizushima, S. (1990) J. Biol. Chem. 265, 21166–21171.
[14] Engelman, D.M. and Steitz, T.A. (1981) Cell 23, 411–422.
[15] Westhoff, P., Farchaus, J.W. and Herrmann, R.G. (1986) Curr. Genet. 11, 165–169.
[16] Von Heijne, G., Steppuhn, J. and Herrmann, G. (1989) Eur. J. Biochem. 180, 535–545.
[17] Hageman, J., Robinson, C., Smeekens, S. and Weisbeek, P. (1986) Nature 324, 567–569.
[18] Kirwin, P.M., Elderfield, P.D. and Robinson, C. (1987) J. Biol. Chem. 262, 16386–16390.
[19] Kirwin, P.M., Elderfield, P.D., Williams, R.S. and Robinson, C. (1988) J. Biol. Chem. 263, 18128–18132.
[20] Von Heijne, G. (1988) Biochim. Biophys. Acta 947, 307–333.
[21] Hudson, G.S., Mason, J.G., Holton, T.A., Koller, B., Cox, G.B., Whitfeld, P.R. and Bottomley, W. (1987) J. Mol. Biol. 196, 283–298.
[22] Deckers-Hebestreit, G. and Altendorf, K. (1986) Eur. J. Biochem. 161, 225–231.
[23] Cashmore, A.R. (1984) Proc. Natl. Acad. Sci. USA 81, 2960–2964.
[24] Kohorn, B.D. and Tobin, E.M. (1989) Plant Cell 1, 159–166.
[25] Stayton, M.M., Black, M., Bedbrook, J. and Dunsmuir, P. (1986) Nucleic Acids Res. 14, 9781–9796.
[26] Mullet, J.E. (1983) J. Biol. Chem. 258, 9941–9948.
[27] Coruzzi, G., Broglie, R., Cashmore, A. and Chua, N. (1983) J. Biol. Chem. 258, 1399–1402.
[28] Darr, S.C., Somerville, S.C. and Arntzen, C.J. (1986) J. Cell Biol. 103, 733–740.
[29] Bürgi, R., Suter, F. and Zuber, H. (1987) Biochim. Biophys. Acta 890, 346–351.
[30] Berzborn, R.J., Klein-Hitpass, L., Otto, J., Schünemann, S. and Oworah-Nkruma, R. (1990) Z. Naturforsch. 45C, 772–784.
[31] Michel, H.P. and Bennett, J. (1987) FEBS Lett. 212, 103–108.
[32] Michel, H., Hunt, D.F., Shabanowitz, J. and Bennett, J. (1988) J. Biol. Chem. 263, 1123–1130.
[33] Alt, J., Winter, P., Sebald, W., Moser, J.G., Schedel, R., Westhoff, P. and Herrmann, R.G. (1983) Curr. Genet. 7, 129–138.
[34] Sebald, W. and Wachter, E. (1980) FEBS Lett. 122, 307–311.
[35] Tae, G., Black, M.T., Cramer, W.A., Vallon, O. and Bogorad, L. (1988) Biochemistry 27, 9075–9080.
[36] Herrmann, R.G., Alt, J., Schiller, B., Widger, W.R. and Cramer, W.A. (1984) FEBS Lett. 176, 239–244.
[37] Widger, W.R., Cramer, W.A., Hermodson, M., Meyer, D. and Gullifor, M. (1984) J. Biol. Chem. 259, 3870–3876.
[38] Vallon, O., Tae, G., Cramer, W.A., Simpson, D., Hoyer-Hansen, G. and Bogorad, L. (1989) Biochim. Biophys. Acta 975, 132–141.
[39] Heinemeyer, W., Alt, J. and Herrmann, G. (1984) Curr. Genet. 8, 543–549.
[40] Szczepaniak, A. and Cramer, W.A. (1990) J. Biol. Chem. 265, 17720–17726.
[41] Sayre, R.T., Andersson, B. and Bogorad, L. (1986) Cell 47, 601–608.
[42] Zurawski, G., Bohnert, H.J., Whitfeld, P.R. and Bottomley, W. (1982) Proc. Natl. Acad. Sci. USA 79, 7699–7703.
[43] Alt, J., Morris, J., Westhoff, P. and Herrmann, R.G. (1984) Curr. Genet. 8, 597–606.
[44] Willey, D.L., Auffret, A.D. and Gray, J.C. (1984) Cell 36, 555–562.
[45] Szczepaniak, A., Black, M.T. and Cramer, W.A. (1989) Z. Naturforsch. Sect. C. Biosci. 44, 453–461.
[46] Hennig, J. and Herrmann, R.G. (1986) Mol. Gen. Genet. 203, 117–128.
[47] Bird, C.R., Koller, B., Auffret, A.D., Huttly, A.K., Howe, C.J., Dyer, T.A. and Gray, J.C. (1985) EMBO J. 4, 1381–1388.
[48] Franzén, L.-G., Frank, G., Zuber, H. and Rochaix, J.-D. (1989) Mol. Gen. Genet. 219, 137–144.
[49] Widger, W.R., Cramer, W.A., Hermodson, M. and Herrmann, R.G. (1985) FEBS Lett. 191, 186–190.